



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 17, 2017 – 10:05 PM EST

PDB ID : 5U33  
Title : Crystal structure of AacC2c1-sgRNA-extended non-target DNA ternary complex  
Authors : Yang, H.; Gao, P.; Rajashankar, K.R.; Patel, D.J.  
Deposited on : 2016-12-01  
Resolution : 3.75 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028442
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028442

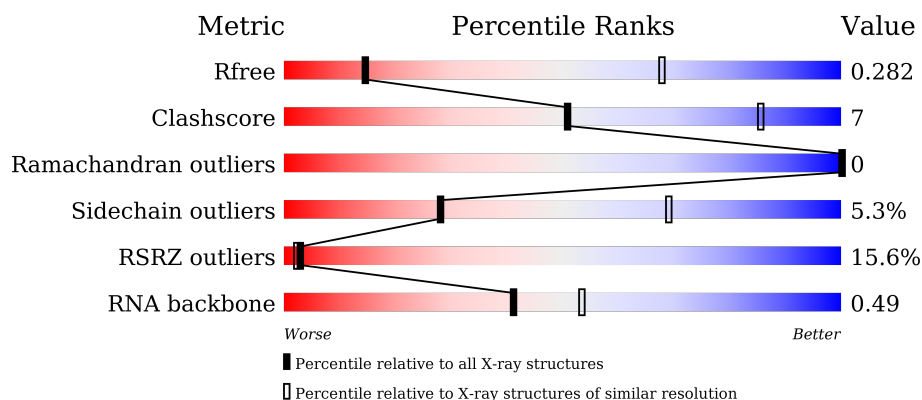
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.75 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.



Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	1268 (4.02-3.50)
Clashscore	102246	1407 (4.02-3.50)
Ramachandran outliers	100387	1346 (4.02-3.50)
Sidechain outliers	100360	1342 (4.02-3.50)
RSRZ outliers	91569	1276 (4.02-3.50)
RNA backbone	2183	1069 (4.70-2.80)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions  $\leq 5\%$ . The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	1130	<div> <div>16%</div> <div> <div></div> <div>78%</div> <div>17%</div> <div>• •</div> </div> </div>
2	B	112	<div> <div>4%</div> <div> <div></div> <div>47%</div> <div>27%</div> <div>16%</div> <div>10%</div> </div> </div>
3	C	28	<div> <div></div> <div> <div></div> <div>71%</div> <div>29%</div> </div> </div>
4	D	28	<div> <div>4%</div> <div> <div></div> <div>39%</div> <div>11%</div> <div>50%</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
5	SO4	A	1201	-	-	-	X

## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 11883 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called CRISPR-associated endonuclease C2c1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1085	Total	C	N	O	S	0	0	0
			8864	5569	1653	1611	31			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	SER	-	expression tag	UNP T0D7A2
A	570	ALA	ASP	engineered mutation	UNP T0D7A2
A	848	ALA	GLU	engineered mutation	UNP T0D7A2
A	977	ALA	ASP	engineered mutation	UNP T0D7A2

- Molecule 2 is a RNA chain called sgRNA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	101	Total	C	N	O	P	0	0	0
			2167	967	401	698	101			

- Molecule 3 is a DNA chain called Target DNA strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	28	Total	C	N	O	P	0	0	0
			565	271	98	169	27			

- Molecule 4 is a DNA chain called Non-target DNA strand.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	14	Total	C	N	O	P	0	0	0
			282	139	38	92	13			

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	O	S	0	0
			5	4	1		



Category	Count
T1	100
G5	100
C8	100
T9	100
T102	100
T103	100
T104	100
DT	100

## 4 Data and refinement statistics

Property	Value	Source
Space group	I 2 2 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	113.50Å 179.11Å 216.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.49 – 3.75 100.54 – 3.75	Depositor EDS
% Data completeness (in resolution range)	95.4 (47.49-3.75) 95.2 (100.54-3.75)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.35 (at 3.78Å)	Xtriage
Refinement program	PHENIX (1.11.1 _2575: ???)	Depositor
R, $R_{free}$	0.245 , 0.286 0.243 , 0.282	Depositor DCC
$R_{free}$ test set	1121 reflections (5.11%)	DCC
Wilson B-factor (Å <sup>2</sup> )	134.0	Xtriage
Anisotropy	0.097	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 101.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.44$ , $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	11883	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	148.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.70% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.



## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 5$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	$\# Z  > 5$	RMSZ	$\# Z  > 5$
1	A	0.26	1/9057 (0.0%)	0.40	0/12202
2	B	0.22	0/2424	0.82	0/3774
3	C	0.51	0/631	0.94	0/971
4	D	0.55	0/311	1.15	0/477
All	All	0.28	1/12423 (0.0%)	0.59	0/17424

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	58	GLU	CD-OE2	6.99	1.33	1.25

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts [i](#)

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	8864	0	8763	125	0
2	B	2167	0	1096	33	0
3	C	565	0	318	5	0
4	D	282	0	167	5	0
5	A	5	0	0	0	0
All	All	11883	0	10344	146	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including hydrogen atoms). The all-atom clashscore for this structure is 7.

The worst 5 of 146 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:130:ASP:HB3	1:A:133:ALA:HB2	1.59	0.84
1:A:1085:ILE:HD13	1:A:1100:VAL:HG22	1.70	0.74
1:A:738:ARG:HB3	2:B:37:G:H5''	1.70	0.73
1:A:761:ILE:HD13	1:A:805:LYS:HG2	1.72	0.71
1:A:418:ILE:HD13	1:A:502:LEU:HD23	1.73	0.71

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	1077/1130 (95%)	1021 (95%)	56 (5%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	930/966 (96%)	881 (95%)	49 (5%)	28	69

5 of 49 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	597	ARG
1	A	652	LEU
1	A	1007	ASP
1	A	607	ASN
1	A	711	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	26	HIS

### 5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
2	B	99/112 (88%)	28 (28%)	5 (5%)

5 of 28 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
2	B	5	A
2	B	6	G
2	B	12	A
2	B	13	G
2	B	16	U

All (5) RNA pucker outliers are listed below:

Mol	Chain	Res	Type
2	B	5	A
2	B	22	A
2	B	32	C
2	B	82	G
2	B	86	U

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
5	SO4	A	1201	-	4,4,4	0.24	0	6,6,6	0.07	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	SO4	A	1201	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 5.7 Other polymers

There are no such residues in this entry.

## 5.8 Polymer linkage issues

The following chains have linkage breaks:

Mol	Chain	Number of breaks
4	D	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	D	10:DT	O3'	101:DT	P	48.98

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<sup>th</sup> percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled ‘Q< 0.9’ lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å <sup>2</sup> )	Q<0.9
1	A	1085/1130 (96%)	0.96	185 (17%) <b>2</b> <b>2</b>	87, 138, 195, 230	0
2	B	101/112 (90%)	0.28	5 (4%) 32 21	100, 156, 257, 271	0
3	C	28/28 (100%)	-0.37	0 <b>100</b> <b>100</b>	98, 130, 149, 161	0
4	D	14/28 (50%)	0.25	1 (7%) 19 11	122, 133, 191, 216	0
All	All	1228/1298 (94%)	0.86	191 (15%) <b>3</b> <b>2</b>	87, 139, 202, 271	0

The worst 5 of 191 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	492	HIS	11.9
1	A	495	ARG	8.7
1	A	494	ARG	7.7
1	A	155	GLY	7.4
1	A	493	ARG	7.3

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors

of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
5	SO4	A	1201	5/5	0.76	0.44	-0.34	143,143,149,152	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.